EAC-CPF Tag Library and Best Practice Guide

Proposal for discussion at the Berlin Meeting, March 2020

Introduction

As part of the redevelopment of EAC-CPF, and in response to feedback from users about usability, there is a need to update the EAC-CPF Tag Library and provide further documentation to assist users with implementing it.

The following paper works with two assumptions: the EAD3 and EAC-CPF tag libraries use the same transformations so the existing structure of the tag library will be retained; and considering the ongoing work towards a shared schema, where possible the tag libraries for the two schemas should be aligned.

This paper is intended to be a starting point for discussions and any proposed solutions are open to changes.

1: "Availability information in the tag library is confusing"
(see: https://github.com/SAA-SDT/eac-cpf-schema/issues/20)

Feedback received indicates that the current EAC-CPF Tag Library does not make it clear which elements are mandatory within a parent element. This issue can be broken up into two issues: the availability of specific elements within parent elements (listed under 'May occur within:'), and the availability of child elements within elements (listed under 'May contain:').

The use of the headings 'May contain' and 'May occur within' implies that all of the listed elements are optional, however this is often not the case. It would be difficult to clearly document which elements in these lists are mandatory because in some cases there is added complexity due to a choice of mandatory elements being available, for example <useDates> must contain one of <date>, <dateRange> or <dateSet>.

1.1 Availability – 'May occur within'

The existing EAC-CPF Tag Library gives a single 'Availability' statement about the conditions and occurrence of an element (exceptions: <citation>, <cpfDescription>, <item>). This does not always reflect the complexity of the availability of elements within different parent elements. For example, the Tag Library lists the availability of <p> as 'Optional, Repeatable', however <descriptiveNote> must contain <p>, so in certain contexts <p> is mandatory. In the case of <dateRange>, the availability statement is 'Optional, Repeatable', which does not accurately describe the complexity of the availability of the element depending on which parent element it is being used within.

The EAD3 Tag Library includes more granular availability statement, listing the specific availability for different parent elements when relevant. EAD3 examples:

<p>

Availability:

- Within <descriptivenote>: Required, repeatable
- Within all other parents: Optional, repeatable

<daterange>

Availability:

- Within <chronitem> and <unitdatestructured>: One of <daterange>, <dateset>, or <datesingle> is required, not repeatable
- Within <dateset>: One of <daterange> or <datesingle> is required, repeatable
- Within <localcontrol> and <relation>: Optional, not repeatable

Proposal: update availability statements to fully document the availability of elements in all contexts.

Example: <dateRange>:

Availability:

- Within <cpfRelation>, <function>, <functionRelation>, <legalStatus>, <localDescription>, <mandate>, <occupation>, <place>, <resourceRelation>: one of <date>, <dateRange> or <dateSet> optional, non-repeatable
- Within <localControl>: one of <dateRange> or <date> optional, non-repeatable
- Within <existDates>, <useDates>: one of <date>, <dateRange> or <dateSet> required, non-repeatable
- Within <chronItem>: one of <date> or <dateRange> required, non-repeatable
- Within <dateSet>: one of <date> or <dateRange> required, repeatable

1.2 Availability – 'May contain'

The EAC-CPF Tag Library lists child elements that an element 'May contain', but it does not clarify which child elements are optional and which are mandatory. Using <descriptiveNote> as an example again, the EAC-CPF Tag Library says it 'May contain: p', however one or more <p> is mandatory within <descriptiveNote>.

EAD3 includes a statement in 'Description and Usage' to detail mandatory child elements. It states that <descriptive note> 'must contain one or more <p> elements'. In some cases EAC-CPF does imply that child elements are required, for example the fourth paragraph on the <nameEntry> description says '*Each <nameEntry> should contain at least one <part> element*', however it could be improved by including a clear and consistent statement early in the description. For example, <persname> in EAD3 starts the second paragraph in the description with '*<persname> must contain one or more <part> elements*'.

Proposal: Include a statement early in the 'Description and usage' section detailing any mandatory child elements using consistent terminology, eg <element1> must contain one or more <element2>.

Example: <nameEntry>

May contain: alternativeForm, authorizedForm, part, preferredForm, useDates

May occur within: identity, nameEntryParallel

Description and Usage:

Each <nameEntry> must contain one or more <part> elements. Within <nameEntry> each of the component parts of a name may be recorded in a separate <part> element.

[…]

2: Consistency of terminology

This is a small issue that I observed while comparing the EAC-CPF and EAD3 tag libraries around other issues - each tag library uses different terminology to describe the same thing. Specifically, under 'Availability', EAC-CPF describes availability being 'Mandatory, Non-repeatable', while EAD3 uses 'Required, Not repeatable'.

Proposal: Adopt common terminology for availability across the EAD and EAC-CPF tag libraries.

3: Consistency of descriptions

As we do further work towards a shared schema for EAC-CPF and EAD, should the description of elements that are identified as having the same name and meaning be unified across both tag libraries? The summaries for elements could be a useful starting point.

Proposal: Work towards using the same text in the 'Summary' for elements in EAC-CPF and EAD where the name and meaning of the element has been identified as being the same.

4: <control>

Concerns have been raised about the way <control> is expressed (https://github.com/SAA-SDT/eac-cpf-schema/issues/23). In the current EAC-CPF Tag Library there are two separate sections that provide information about <control> - in the 'Overview of EAC-CPF Structure and Semantics', and in the description of the <control> element. The 'Overview' provides details of the required and optional sub-elements and the prescribed order of sub-elements which are not included in the separate element details. In EAD3, these details are all included in the <control> element description and there is no overview of the structure.

Proposal: Include more details about sub-elements in the description of the <control> element, particularly the prescribed order and whether they are required or optional.

Also include complete and more complex examples and explanations of <control> in the Best Practice Guide.

Consider: Is it useful to continue to include the 'Overview of EAC-CPF Structure and Semantics' in the Tag Library, either in its entirety or in a simplified form (for example only including the introduction that gives a high-level overview of the structures)? Should this information be integrated into the element descriptions so it is not split across two sections of the Tag Library? Should it be included in the Best Practice Guide instead of the Tag Library?

5. Other Sections in Current Tag Library

In addition to the details of the elements and attributes, the current EAC-CPF Tag Library includes several explanatory sections. One example of this is the 'Overview of EAC-CPF Structure and Semantics' which is discussed in the previous section in relation to the <control> element. There is also a 'Background' section, 'EAC-CPF Concepts' and a 'Glossary.' While these include a lot of useful information, some of this descriptive text might be better included in the new Best Practice Guide

The EAD3 Tag Library includes 'Tag Library Conventions' (https://www.loc.gov/ead/EAD3taglib/EAD3.html#d1e307) where the structure of the element descriptions are explained, which might be useful to include in the EAC-CPF Tag Library.

Consider: Which information needs to be included in the Tag Library? Can some of the more detailed explanatory text be included in the Best Practice Guide instead?

6: Best Practice Guide

There has been discussion of the need to provide more examples of EAC-CPF in use (see for example https://github.com/SAA-SDT/eac-cpf-schema/issues/37). It was agreed at the 2019 EAC-CPF meeting in Austin that we would create a Best Practice Guide as part of, or to accompany, the EAC-CPF Tag Library.

The following is an extract from the 'Topic: Names' paper presented at the Austin meeting that describes the purpose of the best practice guide (see: https://github.com/SAA-SDT/TS-EAS-subteam-notes/blob/master/eaccpf-subteam/working-documents/topics/names/Topic_names_20191220.pdf):

> *A. Add a new chapter Best Practice to the Tag Library / Documentation*
>
> - *Explain the usage of a set of describing elements and attributes with different use cases, e.g. for names, dates, rules, other complex objects.*
> - *Give best practise examples from EAC-CPF users for the different use cases.*
> - *Show example encodings, designed or real ones, to clarify the encoding.*
> - *Provide an ideal example encoding and a short one*

The Best Practice Guide could be a useful place to explain new functionality, such as the assertion descriptions, and provide some detailed examples of its intended use.

The new issues in GitHub for individual elements are already identifying elements that might need to be included in the best practice guide under the 'Solution Documentation' sub-heading (see for example: https://github.com/SAA-SDT/eac-cpf-schema/issues/116).

As this proposal was accepted during the Austin meeting, there is no new proposal here. Rather the Best Practice Guide is being documented so we can consider how to identify elements or groups of elements that need to be included in the Guide, and how to gather examples.

7. Strategy for capturing Tag Library Content

Proposal: Use the GitHub issues that have been created for each element and attribute to capture Tag Library details, such as new or revised 'Summary', 'Description and Usage' and 'Examples' as the elements are finalised.

Silke has already started identifying elements as 'Topic for Best Practice Guide' and 'New or other example needed'. It would be good to continue with this convention for identifying elements in these categories and work through the elements that need updating.

It would be good to include examples from a variety of different implementations of EAC-CPF. For some of the new or significantly revised functionality examples may need to be developed. We can seek examples from the EAC-CPF sub-team, the broader TS-EAS team, and potentially from other users via the EAD email list.